Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 3 (2011) 556–562

Procedia Computer Science

www.elsevier.com/locate/procedia

WCIT-2010

Secure web-based communication

Nighat Mir [a], Sayed Afaq Hussain [b]

[a] *Lecturer, nighatmir@gmail.com, College of Engineering-Computer Science Department, Effat University, Jeddah, Saudi Arabia*
[b] *Professor, drafaqh@gmail.com, Faculty of Engineering and Computer Science, Riphah International University, Islamabad, Pakistan*

Abstract

With the increase in Internet Technologies, great amount of information is following electronically everyday over the network. Information security is a way to protect information against its confidentiality, reliability and availability.
Hiding exchange of information is an important factor in the field of security. Cryptography and Steganography are two very important methods for this purpose and are both used to ensure data confidentiality. In Steganography a cover media is used to hide the existence of data where cryptography is used to protect information by transferring plain text into cipher text.
Different methods have been studied for multimedia objects but there are very few methods for hiding information into text without altering its integrity. Web based attacks have been a very common practice in recent years and hence need strong security mechanisms for the sake of secret communication. Many robust algorithms can be developed using text Steganography for web pages as they contain a wide amount of bandwidth. A few techniques using web tools like HTML and XML have been proposed but they do not make use of features of these languages very well. This paper discusses some proposed methods, implementations of different embedding techniques and two different ways for hiding data and also a comparative analysis is made based upon some security variables. Text Steganography is applied on XML files and is further encrypted using a cryptographic algorithm.

Keywords: Text Steganography;cryptogrpahy; Extensible Markup Language (XML); Advanced Encryption Standard (AES); secure web; Algorithm;

1. Introduction:

Secret communication has been a subject of interest for ages. With the vast expansion of Internet, massive web based information is travelling every day and securing data is a very important subject in this matter. For security reasoning, many different methods have been implemented and new methods are evolving every day. Cryptography, Steganography and Watermarking are well known ways of securing information but they all work under different mechanisms.

Cryptography makes data unreadable by writing into secret code and it ensures authentication, confidentiality and integrity [1]. Steganography hides the existence of data and it ensures transparency, robustness and capacity. Whereas, watermarking technique provides evidence for the intellectual property rights over certain content by hiding some information in it.

Web based communication has a great amount of bandwidth and hence can be used for secret communication. HTML and XML are two basic but important and universal tools for web development. Scripting languages are used for dynamic web development but all browsers at the end have to translate the scripting code into HTML format.

doi:10.1016/j.procs.2010.12.092

Steganography has mostly been applied on multimedia and voice based information but very few methods have been developed for the text documents. Text Steganography is an extension of Steganography. A cover medium is required to hide information in the same way. In this paper some Text Stegagnographic methods are applied on XML file and is further combined with the cryptography to add another layer of security. XML data is used as a cover medium and AES algorithm is used to secure it further. Combining Cryptography with Steganography has been applied in different ways but for multimedia based information [2] but problem of less research has so far been noticed in the field of Text Steganography. XML is considered as a medium because of its flexibility by creating our own tags which could be exploited to hide information in a better way, since there is less chance of vulnerability of a document.

1.1. Related Work:

Shingo and Kyoko in [3] have proposed some techniques for hiding information using XML file. Techniques discussed were empty elements, white spaces in tags, attribute and element ordering. All these techniques have been proposed as a communication model; none of these were implemented to show the results of these techniques but proposed as a future recommendation. Experiments and analysis prove that different ways can be adopted to use each method hence varieties are noticed amongst different authors. Some more techniques have also been proposed in [4] and [5] but only for HTML files. Mohammed and Sun in [4] have proposed some watermarking techniques for HTML based pages. They have focused on exploiting white space, line breaks, attributes ordering, string delimiter and color values. All above mentioned techniques were not implemented however, some of these have been tested to show sample results. In [5] by Ala'a and Mazin have also used HTML files to achieve secret communication. Their idea was to hide a secret data in an HTML file by using white space inside the webpage text and then to encrypt only colored data by using DES algorithm. Capacity of hidden information is not enough in this case as only white spaces were considered for hiding textual information. Aasma, Sumbul and Asadullah in [6] have also studied information hiding methods using XML files. They have tested random character, reverse character, tags shuffling and attribute shuffling techniques and have shown results with respect to the security and bandwidth.

1.2. Proposed Methodology:

XML files are used thoroughly in this system for achieving a secret communication for textual information only. On the sender side, XML file is used as a carrier with cover data which is further encrypted using AES (Advanced Encryption Standard) algorithm, where text information is converted into a string message. This string message is embedded into XML file using nine different embedding techniques which are described in section 2.1 accordingly. All embedding techniques have therefore been implemented to see the actual working scenario. A reverse procedure is applied on the receiver side to get back the original file by first decrypting and then the original XML where embedded data is extracted.

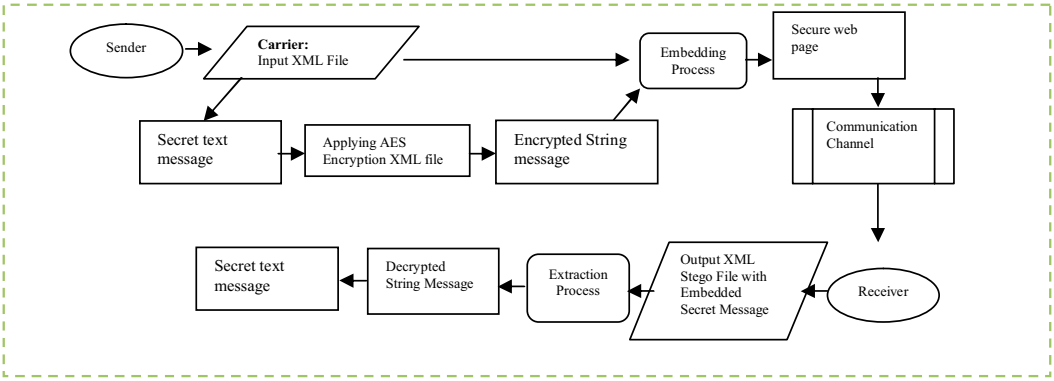

Figure 1: Flow Chart of System

2. Algorithm:

XML file is taken as an input on sender side which is embedded using one of the nine implemented embedding methods and is further, encrypted using a cryptographic algorithm-AES (Advanced Encryption Standard).

A reverse procedure is applied on stegoed-encrypted XML file which will first be decrypted using AES and then extracted to get back the original file.

2.1: Embedding Techniques:

i. White Space Method:

Data is embedded by inserting a white space in XML tags. In this method an extra space is inserted in alignment with tags. A space is inserted after reading "<" and before reading ">" character. White space represents a one-bit data in XML file. A reverse procedure is applied in extraction phase when all extra spaces in tags are removed.

ii. White Space Replacement Method:

Data is embedded by replacing all white or extra spaces in XML file with "&nbps" character value. A reverse method of replacing the "&nbps" character value with empty space is applied to extract the original message.

iii. Empty Tags Method:

Data is embedded by using empty tags. Representation of an empty tag is either a start-tag immediately followed by an end tag or an empty tag. Usually this technique can be exploited or implemented using the <img> tag. In this method the first image tag is taken and a closing character "/" is inserted before reading the end character ">".

We need to have both <tag/> and <tag> </tag> to carry out the extraction process. So that when "/" character is erased from the first tag <tag/>, it will have another closing tag </tag> to avoid any error. In reverse process "/" is deleted before reading the end character ">".

iv. Random Characters Method:

Data is embedded by inserting random characters in tags. A character is inserted after reading the first character of first tag. Similarly, after each word one random character is inserted. In case of a special character, full stop (.), wild character or white space the process is repeated again. Embedding process is recursively applied to all tags. At the reverse process, all inserted characters are deleted from the file.

v. Color Replacement Method:

Data is embedded by replacing color name with its hexadecimal value. In this method it first needs to find the color attribute followed by character "=" and then the color name is replaced with its hexadecimal value. The process is recursively applied on all color tags in XML file. At reverse, the hexadecimal value is replaced back to its color name.

vi. Line Break Method:

Data is embedded by exploiting Line Break. In this method, first tag is taken and a line break is inserted after reading the closing character ">". The process is recursively applied to all tags. A reverse process is applied to erase all line breaks to get back the original format.

vii. Word Space Method:

Data is embedded using tag. In the method Text is first divided into blocks of word. A data bit is then embedded by adjusting the width of spaces between the characters within a Block, according to some predefined rule. A block size of 18-20 characters is predefined and within this block size 3-bits data (secret message) is embedded. At reverse extra spaces are removed to get back the original file.

viii. Synonyms Method:

Data is embedded by replacing words with their synonyms. In this method, first a Synonyms Word List (SWL) is created as a database and XML file is compared with this list. Every word from the text <body> tag of XML file is compared with the list and if a match is found, it is replaced by its synonyms otherwise is ignored. This process is recursively applied till the end of file. At reverse, one needs to have a list of synonyms to extract the original message.

ix. Acronyms Method:

Data is embedded by exploiting Acronyms to hide data. In this method first an Acronyms Word List (AWL) is created based upon standard acronyms. An acronym is selected to hide and exploit the hidden message. XML file is created using the altered acronyms. At reverse, one needs to have a list of acronyms to extract the original message.

3. Results:

i. ***White Space Method:***

Original data, without any space: <tag> information</tag>**;** *Embedded data, after inserting space* <tag > information</tag >.

Hidden message is read as a string of 1's and 0's, every extra space corresponds to 1 and none with 0, hence a series of 1's and 0's were extracted to generate a message.

ii. ***White Space Replacement Method:***

Original data, without any space: <tag> information </tag>; *Embedded data, after inserting space* <tag>&nbps;text&nbps;text</tag >.

Hidden message is read as a string of 1's and 0's, every "&nbps" corresponds to 1 and none with 0. All space values were extracted to read the actual message.

iii. ***Empty Tags Method:***

Original data, without using empty tags: <img> imgname </img>; *Embedded data, using empty tags*
<tag information />

Only <img> tag was considered in this method, and "/" character with <img> tag was considered representing a value 1.

iv. ***Random Characters Method:***

Original data, without inserting random characters: <tag> information </tag>; *Embedded data, inserting random characters* <tag> infaormmettihoins</tag *(namethis is an embedded message in this tag).*

All extra characters correspond to 1's and hence a string of 1's and 0's was extracted to read the extra information.

v. ***Color Replacement Method***

Original data, without replacing colour name: <text colour = "blue">; *Embedded data, after replacing colour value* <text color = "#0000FF">

This method was tested only using 6 basic colours. A colour list is required, and only mentioned colours were replaced with their hexadecimal values and rest were not considered.

vi. ***Line Break Method***

Original data, without line break; text ; E*mbedded data, after inserting line break*

text
Line break with any tag was considered as a value 1 and normal tags with 0, to generate a sequence of message.

vii. ***Word Space Method***

Original data, without word space; The primary aim of the Faculty Professional Development Committee is to promote professional an academic development within the three departments of the College of Engineering in line with the College of Engineering and the University's strategic goals.

Embedded data, after inserting word space
Block1: The primary aim of the
Block2: Faculty Professional
Block 3: Development Committee
Block4: Committee is to promote
Block 5: professional an acade
This method was tested only on tag for textual information, where a block size was defined to insert an extra space between words and hence a series of 1's and 0's was extracted from the blocks of words.

viii. ***Synonyms Method:***

Original data, without replacing words;
"Steganography has long been used to facilitate secret communications Tattooed heads concealed by hair covered wooden tables with wax and even digital watermarking in the modern age have been used to conceal message inside another to send secret notes in enclosed messages"
Embedded data, after replacing words with their synonyms
Secret message: "mail enclosed note"
SWL:

Words (cover text)	Synonyms
Concealed	masked
Facilitate	help
Head	top
Covered	enclosed
Message	note
Modern	recent
Send	mail

Table 1: List of Synonyms

A synonym list is required is method, where every word from <text> tag is matched with the list. When a match is found a word is replaced with its synonym

ix. ***Acronyms Method:***

Acronym list used to embed data

Words (cover text)	Acronyms	Translated
B2B	Business to Business	Back to Boat
ACK	Acknowledgement	Air coach kids
ALT	Alternate	Add list towards
BCC	Blind Carbon Copy	Blue camp car

Table 2: List of Acronyms

Acronym list with predefined translated acronyms has been defined wherein acronyms used in file is translated according to the list. A person having the AWL can only read the hidden message and there is no chance to get the

message. Acronyms used in the file are standards together with their universal meanings. Acronyms used in the file correspond to 1's and the rest to 0's.

4. Comparative Analysis:

Embedding techniques have been analyzed with respect to different variables (i.e. Blindness, Security, capacity and robustness). Limitations of each method have also been analyzed and discussed in accordance with comparative analysis. White space method has been considered robust, un-noticeable, secure and of medium capacity. It doesn't have good capacity as too much white space on any web page may disturb the appearance of rich information and also can make it vulnerable. White space replacement has been considered noticeable, as replaced character "&nbps" may look odd in coding if other information in page also not converted in different formats, like colours not converted into hexadecimal etc. capacity is medium like the first one as, a web page can sustain white spaces to a certain extent. Empty tags method has been considered robust, un-noticeable and secure but with low capacity, as this method has been applied only on certain tags. Random characters method has a high capacity as a lot of text appears on web pages but is noticeable so does not satisfy security feature. In some cases it may disturb the actual meaning of text with wrong or extra words. This method can be improved by applying some more parameters on it. Colour replacement method has been considered un-noticeable, secure but has low capacity in our case, as it has a limitation of having a colour list for some specific colours. Capacity can be improved to high by adding more colours in list. Word space method has been considered un-noticeable, secure and has medium capacity. Synonyms method has been considered un-noticeable, secure and of high capacity but have limitations of having a list of synonyms for matching. Acronyms method has been considered un-noticeable, secure and of high capacity but this also requires a list of acronyms for translating the words. Both synonyms and acronyms would be considered secure in case when lists are in hands of user but their security is breached if list is passed to a third party.

5. Conclusion and Future Work:

In this paper, we have presented Text Steganography combined with Cryptography for hiding secret information using XML files. Nine different embedding techniques studied and applied on XML file. System has been implemented using C# language for all nine methods combined with AES which has added another layer of security. All methods are measured with respect to different standards and it is analyzed that white space method, white space replacement method, colour replacement method, line break method synonyms method and acronyms methods are considered stronger and less vulnerable. Furthermore, improvements in colour, synonyms and acronyms are needed to make them more practical, efficient and stronger.

Techniques discussed in this paper have therefore been applied on textual information and hence could also be applied on other types of data in XML files, as XML does not only contain text but multimedia based information as well and the idea could be extended toward other parts. These Embedding techniques can also be applied on other web tools like scripting languages. Moreover, a bandwidth or capacity comparison between mark-up languages and scripting languages can also be obtained.

References

1. S. Al-Riyami, K Paterson, "Advances in Cryptology-ASIACRYPT" 2003, Springer.
2. Syed Afaq H., M. Sikander, Nighat Mir, Beenish, "A Secure Model for Data Communication using Cryptography and Steganography" 2006, ICT Malaysia.
3. Shingo, Kyoko, Ichiro, Osamu, "A Proposal on Information Hiding Methods using XML", Mitsubishi Research Institute, Communication Research Laboratory, Yokohama National University and The University of Tokyo.
4. Mohammad Laheen, Sun XingMing, "Techniques with Statistics for Web page Watermarking" 2005, NSFC No.60373062.
5. Ala`a H., Mazin S., Mohammad A. Al Hamami, "A Proposed Method to Hide inside HTML Web Page File" .
6. Aasma, Sumbul, Asadullah, "Steganography: A New Horizon for Safe Communication through XML", 2005-

2008, JATIT